STAT 5120: Applied Linear Models
Fall 2017

Instructor: Chao DU
E-mail: cd2wb@virginia.edu
Phone: 434-924-3014
Office: Halsey 107

Class Schedule: TuTh 5:00pm-6:15pm
Location: Claude Moore Nursing Educ G120
Office Hour: Thursday 12pm-2pm
Teaching Fellow: Ye Lin yl5kn@virginia.edu

Course Description:

This course focuses on the so-called linear regression analysis, one of the oldest, and the most fundamental approach in statistics science. Broadly speaking, the term “linear regression analysis” refers to a large number of different statistical methods that focus on analyzing the relationship between dependent variables and linear function of independent variables. The history of linear regression analysis can be traced back to Legendre and Gauss at the beginning of 19th century, who developed a method named “Least Squares” to determine the orbits of astronomical bodies about the Sun. Since then, the idea of “linear regression model” has been widely used in various areas, from determining the relationships between body weight and height to investigating how development of a city could depend on different factors such as incomes, temperature, education and ethnicity. And in the age of data science, the technique of linear regression still remains as one of the most powerful weapon in the arsenal of data scientists.

This course is designed for statistics-majored undergraduates and non-statistics-majored graduate students. The aims of this course are: 1) to provide a broad overview of the linear regression methods that are available to data analysts and researchers; 2) to help students develop critical thinking skills for understanding the structure and information contained in data and interpreting the inference results; 3) to master necessary computing skills for analyzing data using linear regression model; 4) to introduce essential theoretical foundation that serves as a starting point for future work. While this course is applied in nature and involves a large number of data examples, it will also cover in-depth discussion on the motivations, as well as the mathematical foundations of different linear regression methods.

Textbook:

Introduction to Linear Regression Analysis, by Montgomery, Peck, and Vining, 5th ed.

The lecture will cover many topics discussed in the textbook but is structured differently.

Prerequisites:
As this course involves a fair amount of theoretical discussion, students enrolled in this class should be comfortable with multivariate calculus and matrix algebra, as well as basic mathematical statistics. It is highly recommended that the students should already have taken the following classes:

STAT 3120, and either MATH 3351 or APMA 3080.

An open-book quiz (not counting towards final grade) will be held on August 24 in class to assess your preparation for this class. If you feel your preparation is weak, please consult with me about your concerns.

**Course Assessment and Relevant Policies:**

The final numerical grades will be calculated based on the major factors described below.

1. **Homework (25%).**
   - There will be around 10 written assignments. Two assignments with lowest scores will be dropped for the purpose of calculating the final grades and the rest will be weighted equally.
   - Assignments will include theoretical questions, short answer questions based on computer output for a data set, and analyzing real data sets using R. A short tutorial will be provided to help with R code.
   - For each assignment, a few problems will be graded based on correctness, while the rest will be graded on completeness.
   - Each assignment must be submitted on time. Missed or late assignments will receive a grade of zero with the exceptions of medical conditions, other unforeseen emergencies and official University activities. See the general class policies section for detail.

2. **Exams (75%, two midterm exams and one final exam).**
   - Each midterm is worth 20% and the final exam is worth 35% for calculating the final grade.
   - All exams will be graded based on correctness.
   - All exams are closed book-exam but you can bring written notes as well as a calculator. In particular, you can bring 1 sheet of notes for the first midterm, 2 sheets for the second midterm, and 3 sheets for the final. The size of each sheet of notes must not exceed the size of a A4-sized paper but you can utilize both sides. As the process of preparing notes can help you gain an overview of the class materials, you must prepare your own notes. PHOTOCOPY and PRINTING are NOT allowed.
• A calculator capable of basic arithmetic operation is needed for the exam. Smart phone is not allowed for this purpose.

• Missed exam will receive a grade of zero with the exceptions of medical conditions, other unforeseen emergencies and official University activities. See the class policy section for detail.

• The schedule of exams is listed below:

<table>
<thead>
<tr>
<th>Exam</th>
<th>Date</th>
<th>Time</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Mid-Exam</td>
<td>Sep 28</td>
<td></td>
<td>in Class</td>
</tr>
<tr>
<td>2nd Mid-Exam</td>
<td>Oct 31</td>
<td></td>
<td>in Class</td>
</tr>
<tr>
<td>Final Exam</td>
<td>Dec 7</td>
<td>7:00pm-10:00pm</td>
<td>location TBD</td>
</tr>
</tbody>
</table>

3. Class participation (Extra Credit).

• You should attend all lectures as it is the most effective way to learn the class materials. You should also take advantage of the regular office hours if you have any questions regarding class materials.

• Regular and active class participation can bump you up one third of a letter grade (such as B to B+) if you are on the “border” of a grade.

**UVaCollab and Class Announcement:**

A workspace for the course will be maintained at UVaCollab. Course materials will generally be posted by the evening before class. Assignments and other materials will also be posted throughout the semester. Announcements regarding the posting of course assignments and other class-related subjects will be sent through emails (the subject line will start with [Stat 5120]). Please make sure to check your email regularly (at least once a day) to keep updated.

**Computing:**

The course includes a fairly extensive practical computing component. R will be used as the “official” statistical computing software in this course. It is free and can be downloaded from [http://cran.r-project.org/](http://cran.r-project.org/). You will not be asked to do complicated programming for this course. However, we do expect that at the end of this course, you will be able to carry out regression analysis on a dataset using R and interpret the relevant R outputs.

We will discuss examples on how to use R in class. And each assignment will also include a tutorial to help you with the assignment. R code for examples during lectures will also be provided. Still, due to the limited time we could spend for each class, you are expected to spend time going over those examples on your own.
The best rehouse to learn R is through the internets. And below you may find some useful reference to start with:

R project homepage:  https://www.r-project.org

Online R Tutorial:  http://www.cyclismo.org/tutorial/R/

An Introduction to R:  https://onlinecourses.science.psu.edu/statprogram/sites/onlinecourses.science.psu.edu.statprogram/files/lesson00/R-intro.pdf

General Class Policies:

Communications with Instructor

• As a general policy, I will NOT answer questions related to the course material (including assignments or any problem-solving questions) by email. If you have questions regarding the course material, please attend my regular office hour or send me email to schedule an appointment. For this reason, I strongly recommend you to start to work on the assignments early and do not leave everything to the last minute.
• You can send other course-related questions to me. However, I can not guarantee a fast reply. If you have any special request that demands a timely reply, please make sure to send it at least 3 days in advance.
• The subject line of any course-related email should be started with [Stat 5120]. I reserve the right to discard emails that do not follow this format.
• You are welcomed to send me your comments and suggestions through email or discuss with me in person. You can also leave anonymous notes in my mailbox or the course website.

Grading

• Any questions about homework grades should be direct to the teaching assistant first. If you have questions about an exam grade, or cannot resolve an issue with the teaching assistant, please visit my office during office hours or send me email to make the arrangement.
• The grader reserves the right to mark off for untidy or unclear submitted work. All work must be written (or typed) neatly, with each problem clearly labeled and final answers clearly indicated. Pages must be stapled, or neatly bound, in the correct order, and you name must appear on the first page. You must show all work for consideration of partial credit.

Special Circumstances:

• If you need course adaptation or accommodations because of disability, medical emergencies, or other circumstances, please notify me to make appropriate arrangement.
• As a general rule, late or missed assignments and exams will receive a grade of zero. However, the following exceptions apply:
  • 1) No penalties are incurred for late homework or missed examinations because of a medical condition or other reason over which the student has no control. Please make sure to notify me to make appropriate arrangements when the situation allows.
  • 2) If you are involved in an official University activity (e.g., out-of-town competition) or an activity directly related to your academic program (e.g., conference) that cannot be scheduled at another time, please notify me at least three weeks prior to the date that the missed homework will be assigned or missed exam is scheduled, so that alternative arrangements can be made.

Honor Policy:

• As the only true way to acquire knowledge is through your own hard work, it is of the uttermost importance that all the submitted works, such as homework assignments and exam papers, must reflect your independent efforts made during the learning process.
• You may discuss homework assignments in small groups, but each student must finish his or her assignments independently based on his or her own understanding. Copying others’ works will not be tolerated. You must not consult any external resource other than the allowed sheets of written notes during midterm and final exams.
• All submitted work shall be subject to the stipulations of the University of Virginia Honor System. For more information refer to http://www.virginia.edu/honor/.

Class Schedule (tentative) and Important Dates

Week 1 (08/20-08/26)
Introduction.

Preparation Quiz on Aug 24.

Week 2 (08/27-09/02)
Simple linear regression, estimation of parameters, confidence interval, hypothesis test and prediction

Week 3 (09/03-09/09)
ANOVA and F-test, regression assumptions, model diagnostic.

Last Day to drop course without Penalty: Sep 06.

Week 4 (09/10-09/16)
Test for lack of fit, transformation, simultaneous inference.
Week 5 (09/17-09/23)
Matrix tools, multiple Linear regression, general linear F-test, sum of squares.

Week 6 (09/24-09/30)
Multicollinearity.

1st midterm exam on Sept 28.

Week 7 (10/01-10/07)
Model with single binary categorical predictor

No class on Oct 03 (Reading Days).

Week 8 (10/08-10/14)
Model with multiple categorical predictors with multiple levels, polynomial regression.

Week 9 (10/15-10/21)
Model comparison and selection methods.

Last Day to drop course: Oct 17.

Week 10 (10/22-10/28)
Outliers and influential points.

Week 11 (10/29-11/04)

2nd midterm exam on Oct 31.

Linearity and constant variance of error assumptions, added-variable plot.

Week 12 (11/05-11/11)

Cases study.

Week 13 (11/12-11/18)

Logistic regression

Week 14 (11/19-11/25)

Poisson regression

No class on November 26th (Thanksgiving Holiday)

Week 15 (11/26-12/02)
Generalized linear models, multinomial logistic regression.

*Week 16 (12/03-12/09)*

Review on Dec 03.

**Final Exam on Dec 07, 7:00pm-10:00pm.**